Link between the Proposed Task and the Theoretical Framework

Here we link the ideal features presented in our framework to features of our example system. We are grouping features by the main principle of dialogic teaching (Alexander, 2010) that they would enforce. Each principle is presented together with the associated indicators and oracy skills. Below each set of principles, indicators and oracy skills, we comment on the system features related to those elements of pedagogical theory. Each feature from our framework is presented in a bullet point, and below it we specify to which extent that feature would be present in our proposed task. It must be borne in mind that, to improve readability, the framework has been summarized and the author references removed. Throughout this section we mention by number the proposed subtasks described in more detail in a separate document; for clarity, we summarize them below.

Task summary:

1. Introduction and guidance
2. Text reading
3. Paragraph identification (thesis and arguments)
4. Thesis and argument rephrasing
5. Scaffolded discussion

Feature set 1: Cumulative principle

Principle: Cumulative.
Indicator: Uptake.
Oracy skills: Cognitive, Social and emotional, Linguistic.

- Keeping track of which arguments have already been discussed.
    - To be implemented for ST5.
- Displaying chat logs.
    - To be implemented, allowing the student to scroll through the whole conversation and possibly to save it into a file.
- Generating questions on the topic of the discussion: identifying the topic and retrieving information on the topic
    - Open questions not contemplated, only when asking the student to define their position on the topic. The text used in ST2 through ST4 allows the system to set and thus know the topic; datasets would be needed for the system to learn to identify subtopics in the students' arguments in ST5 and to retrieve relevant counterarguments.
- Stance detection
    - To be implemented for ST5.
- Encouraging the student to use their own words.
    - To be implemented in ST4.
- Introducing appropriate terminology.
    - To be presented by the text used in ST2 through ST4. Links to Wikipedia or other reference sites could be included to help students understand the terminology.

Feature set 2: Collective principle

Principle: Collective.
Indicator: Open discussion.
Oracy skill: Social and emotional.

- Measuring the length of the exchanges.
    - To be implemented in ST2 (to see if the student read the whole text), ST4 (to see if the student focused on the text parts relevant to the current stage of the task or if they instead summarized the whole text), and possibly in ST5 if the system had trouble analyzing lengthy arguments.
- Presenting a wide variety of ideas on the topic.
    - To be implemented to some extent in ST5 by introducing counterarguments.
- Matching the topic of the system's and the student's utterances.
    - The topic would be set by the text used in ST2 through ST4, allowing the system not to stray from it; the input from the student would be analyzed to detect when it diverged too much from the topic, and the student would be encouraged to return to the topic.

Feature set 3: Purposeful principle

Principle: Purposeful.
Indicator: Student reasoning.
Oracy skill: Cognitive.

- Setting the goal of the exchange and showing the student what can be achieved through the discussion.
    - To be implemented in ST1, where the system would tell the student the teaching objectives of the task, as well as its more tangible products (a diagram that could be used to develop a coherent argumentation).
- Providing background information.
    - To be implemented in ST5 when the student could not answer; sites such as Idebate[16] could be used to obtain this information.
- Access to chat logs.
    - To be implemented, allowing the student to scroll through the whole conversation and possibly to save it unto a file.
- Frequent feedback that motivates students and informs them of their progress.
    - To be implemented in all subtasks that require student input (ST2 through ST5).
- Identifying reasoning in the student's answers.
    - To be implemented in ST4 by analyzing whether the student included at least one argument from the text, and in ST5 possibly by matching the students' arguments with a larger dataset of arguments on the topic.
- Encouraging reasoning.
    - To be implemented in task descriptions (overall description in ST1 and specifically in ST5) and by asking the student to identify and rephrase arguments in ST3 and ST4 and to use their own arguments in ST5.
- Modelling productive dialogue.
    - To be implemented to some extent by using a model text in ST2 through ST4; however, the text would be rather monologic.

Feature set 4: Reciprocal principle

Principle: Reciprocal.
Indicator: Student questions, Open discussion, High-order questions.
Oracy skill: Cognitive, Social and emotional.

- Countering the student's arguments or reacting with questions.
    - Open questions not contemplated, only when asking the student to define their position on the topic. However, counterargumentation would be implemented in ST5.
- Showing openness to questions.
    - To be implemented to some extent in ST1, where the system should inform the student that they can ask for assistance; other types of questions are not contemplated.
- Asking for clarification.
    - To be implemented through the whole task: the system needs to be confident that it understood the student's input to provide adequate feedback what level of understanding would be necessary would need to be assessed experimentally.
- Giving explanations considering the level of understanding that the students have shown.
    - To be implemented to some extent: the texts used as model for ST2 through ST4 were selected to be suitable for high-school students; resources for output generation in part 5 would also have to be adapted to the expected level of the students. Nonetheless, adaptation for individual level differences are not contemplated.
- Mixed-initiative design.
    - Design is system-initiative, as it is the dialogue system which guides the task. A mixed-initiative design would be exceedingly difficult to implement (Jurafsky and Martin, 2019) and the effort might not be warranted in a task more focused on scaffolding the student through a discussion than on having already skilled students lead an advanced discussion. An option which gives students control of their turns while still using a system-initiative system is the one used by Catania et al. (2020), where students had to press a button to mark the beginning and end of their turns. This also reduces the burden on the speech recognition system, as it would otherwise need to be adjusted to recognize utterances by young users, who tend to take longer to start their turns (ibid).
- Helping the students see how much the dialogue system can understand.
    - To be implemented to some extent in ST1, where the system would give the student some hints about how they can communicate with the system.
- Asking “high order” questions.
    - To be implemented to some extent in ST5, where the student is asked to defend a position and then to respond to counterarguments.
- Large and diverse data sources and a deep-learning approach to generate answers that seem natural and contribute to the discussion.
    - To be implemented at least to some extent in ST5: diverse datasets would be needed to return output based on either information retrieval or on deep learning – which design is selected would have to be decided through tests.

Feature set 5: Supportive principle

Principle: Supportive.
Indicator: [Not directly observable].
Oracy skill: Social and emotional, Linguistic, Physical.

- Respectful and encouraging language.
    - To be implemented throughout the task: prewritten answers are to feature a respectful and encouraging style, and data used for output generation or retrieval would have to be reviewed not to include disrespectful language.
- Friendly avatar.
    - The literature on avatar design was beyond the scope of our review; we can thus offer no suggestions on the concrete realization of this feature.
- Detecting confrontational talk and demanding that students express their ideas in a respectful, reasoned manner.
    - To be implemented to some extent: in ST2 through ST5, students are asked to accompany claims with arguments. Our planned experiments contemplate disrespectful language in o_-task student answers; further tests would have to be performed to detect such language in otherwise correct answers.
- Multimodality (voice recognition, speech generation, a static or dynamic avatar).
    - To be implemented to some extent: though our experiments are text-only, the task is designed to feature speech recognition and generation. As mentioned above, no suggestions are offered regarding avatar design.